The histone code of the fungal genus *Aspergillus* uncovered by evolutionary and proteomic analyses

Xin Zhang^{1,2}, Roberta Noberini³, Tiziana Bonaldi^{3,4}, Jérôme Collemare^{2,*}, Michael F. Seidl^{1,*}

Author affiliations: ¹Theoretical Biology & Bioinformatics Group, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands; ²Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, the Netherlands; ³Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Via Adamello 16, 20139 Milan, Italy; ⁴Department of Oncology and Haematology-Oncology, University of Milano, Via Santa Sofia 9/1, 20122 Milano, Italy

*These authors contributed equally

*Correspondence: J. Collemare (telephone: (+31) 030 21 22 600; email: j.collemare@wi.knaw.nl) and M. F. Seidl (telephone: (+31) 030 253 36 94; email: m.f.seidl@uu.nl)

Abstract

Chemical modifications of DNA and histone proteins impact the organization of chromatin within the nucleus. Changes in these modifications, catalyzed by different chromatin-modifying enzymes, influence chromatin organization, which in turn is thought to impact the spatial and temporal regulation of gene expression. While combinations of different histone modifications, the histone code, have been studied in several model species, we know very little about histone modifications in the fungal genus *Aspergillus*, whose members are generally well-studied due to their importance as models in cell and molecular biology as well as their medical and biotechnological relevance. Here, we used phylogenetic analyses in 94 *Aspergilli* as well as other fungi to uncover the occurrence and evolutionary trajectories of enzymes and protein complexes with roles in chromatin modifications or regulation. We found that these enzymes and complexes are highly conserved in *Aspergilli*, pointing towards a complex repertoire of chromatin modifications. Nevertheless, we also observed few recent gene duplications or losses, highlighting *Aspergillus* species to further study the roles of specific chromatin modifications. SET7 and other components of PRC2 (Polycomb Repressive Complex 2), which is responsible for trimethylation on histone H3 at lysine 27 in many eukaryotes including fungi, are absent in *Aspergilli* as well as in closely related *Penicillium* species, suggesting that these lost the capacity for this histone modification. We corroborated our computational predictions by performing untargeted mass spectrometry analysis of histone post-translational modifications in *Aspergillus nidulans*. This systematic analysis will pave the way for future research into the complexity of the histone code and its functional implications on genome architecture and gene regulation in fungi.

Keywords

Aspergilli, chromatin, DNA methylation, histone modification, H3K27methylation, mass spectrometry

Data Summary

The predicted proteomes used in this study are publicly available at the JGI (Joint Genome Institute) MycoCosm repository (1); the species names and abbreviations are listed in Table S1. To evaluate the completeness of the predicted proteomes and to obtain a species phylogeny, 758 fungal BUSCO (Benchmarking Universal Single-Copy Ortholog) genes were used, and their names are listed in Table S2. The fasta, trimmed alignment, and maximum-likelihood phylogenetic tree files can be found in Supplementary Data 1 and 2 deposited at Zenodo (doi:10.5281/zenodo.5907090), as well as the mass spectrometry results are mentioned in Table S4.

Impact Statement

Uncovering how fungi regulate gene expression and genome organization is an important step to understand how they control biological processes such as growth or biosynthesis of bioactive molecules. Despite the known importance of chromatin modifications in controlling a plethora of biological processes across eukaryotes, they remain particularly poorly understood in filamentous fungi, even in model organisms like *Aspergilli*. Our systematic evolutionary analysis provides a robust framework for the functional analysis of chromatin modifications in *Aspergillus* species and other fungi. Our results do not only implicate candidate enzymes to play a role in new chromatin modifications, but they also point at species that have experienced duplications or losses of genes encoding enzymes for specific chromatin modifications. We therefore expect that this work will set the stage for future research into the complexity of the histone code and its functional implications on gene regulation and genome organization in fungi.

Introduction

Inside eukaryotic cells, the membrane-bound nucleus facilitates the storage and replication of the genetic material as well as the regulation of gene expression (2). Within the nucleus, DNA is arranged as a chromatin fiber formed by repeated nucleosome units consisting of histone protein octamers (two copies of each H2A, H2B, H3, and H4 histones) that are wrapped by 145-147 bp of DNA (3–5). The chromatin structure is regulated by chemical modifications such as methylation of DNA on cytosines (5mC) and adenines (6mA), and different post-translational modifications (PTMs) on the histone proteins, including methylation and acetylation (6–8). The combination of these different modifications, the histone code, influences the recruitment of diverse nucleosome-associated proteins (e.g., DNA binding proteins), which can cause dynamic transitions between the condensed and transcriptionally silent heterochromatin and the accessible and transcriptionally active euchromatin (9–11). Chromatin dynamics is therefore key for organisms to adapt gene expression patterns in response to a variety of developmental and environmental clues.

In the past decades, several enzymes involved in orchestrating chromatin dynamics have been identified and functionally characterized in diverse eukaryotes (3,12,13). These enzymes can be broadly classified into three distinct types: (i) writers that place modifications on nucleosomes or linker DNA; (ii) erasers that remove these modifications; and (iii) readers that first recognize specific modifications and subsequently recruit writers, erasers, or other proteins to the correct genomic location (14–16). Chromatin modifiers that catalyze the same type of chemical modification often harbor functionally and evolutionarily conserved catalytic domains. For instance, DNA methylation (5mC) is catalyzed by different DNA methyltransferases (DNMTs) which contain a DNA methyltransferase domain, while histone mono-, di-, and tri-methylation are nearly exclusively catalyzed by proteins containing a SET (Su (var) 3-9, E(z) and Trithorax) domain (13). Next to these catalytic domains, the enzymes often contain accessory domains, which can contribute to the overall stability of the enzyme or establish interactions with other proteins (17). Furthermore, these multi-domain proteins,

together with other accessory protein subunits, can form multi-protein complexes to find the correct target region within the genome and modify the exact amino acid residue, collectively contributing to the specific functions of these complexes (18,19). For example, the histone methyltransferase complex PRC2 (Polycomb Repressive Complex 2) is responsible for the mono-, di-, and tri-methylation of histone protein H3 on the lysine at position 27 (abbreviated as H3K27me1/2/3; the same abbreviation principle will be applied below for other histone PTMs) (20–23). Evolutionary analyses have suggested that the key enzymes responsible for chromatin modifications have evolved and significantly expanded early in eukaryotic evolution (24). An analysis of 94 gene families with roles in chromatin modifications reveals that 87 out of 94 gene families emerged more than one billion years ago, and 48 were already present in the last eukaryotic common ancestor (24).

Among eukaryotes, the fungal kingdom is remarkably large and diverse, comprising an estimated 1.5–5 million species with various morphologies, life cycles, and ecological niches (25). The most well-studied model fungal species in respect of chromatin modifications is the unicellular yeast *Saccharomyces cerevisiae*, in which 75 histone PTMs on 43 distinct amino acid positions have been reported with histone methylation and acetylation being most abundant (16). In addition to ancient chromatin modifiers, also lineage-specific innovations such as duplications and losses were found in fungi. For example, Dim-2 and RID (Repeat-Induced Point Mutation (RIP)-Defective) are considered to be fungal-specific DNMTs, while DNMT3 that is typically found in animals and plants seems to be absent in fungi (26–28). However, *S. cerevisiae* is not representative of the complete fungal kingdom, and detailed studies on PTMs are needed for other, less studied fungal species.

The fungal genus *Aspergillus* comprises approximately 350 species, including pathogens of humans, animals, and crops, food contaminants, as well as important cell factories for industrial and medical applications (29). Chromatin dynamics influences gene expression in several *Aspergilli* and regulates growth, sexual development, secondary

metabolite biosynthesis, and virulence (16,30–32). Thus far, very few well-known chromatin modifications and their catalytic protein complexes have been identified and studied in few *Aspergillus* species (16). For example, gene knockout of the H3K9 methyltransferase *Clr4* homolog in *Aspergillus fumigatus* reduced histone H3K9me1/3 results in reduction of radial growth and conidia production, as well as delayed conidiation (31). Similarly, CclA, a subunit of the COMPASS (COMplex of Proteins ASSociated with SET1) complex, is essential for H3K4me2/3 in *A. fumigatus*, and the deletion strain showed increased secondary metabolite production and decreased growth (33). The SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex, which is responsible for histone acetylation, has been functionally characterized in *Aspergillus nidulans* (34). Notably, this complex in *A. nidulans* lacks the Sgf11 and Sus1 subunits that have been described in *S. cerevisiae*, suggesting that this complex evolves in distinct trajectories in different fungi (34). Besides these few examples, we so far lack a complete view of the occurrence and evolution of chromatin modifiers in the *Aspergillus* genus. Here, we report the phylogenetic analyses of core enzymes involved in the deposition and removal of histone PTMs, and assess the conservation of all the subunits belonging to chromatin regulator complexes in 94 diverse Aspergilli. We also performed proteomics analyses to corroborate our computational predictions and reveal a broader spectrum of histone modifications.

Material and Methods

Acquisition of predicted proteomes for 109 fungal species

The predicted proteomes of 109 fungal species were retrieved from JGI (Joint Genome Institute) MycoCosm (<https://mycocosm.jgi.doe.gov/mycocosm/home>) (1) on 24th January 2020. These species comprise 94 different Aspergilli species, 13 Ascomycota species from other genera, as well as two Basidiomycota species (*Coprinopsis cinerea* and *Dendrothele bispora*) (Supplementary Table 1).

Phylogenetic analyses of chromatin modifiers

Phylogenetic trees for five major chromatin modifiers were generated using their conserved catalytic domains. Catalytic domain HMMs (Hidden Markov Model) of DNA methyltransferase domain (PF00145), histone methylation SET domain (PF00856), acetyltransferase domain (PF00583), histone acetyltransferase MOZ/SAS FAMILY domain (PF01853), and histone deacetylase domain (PF00850) were downloaded from the Pfam database (<https://pfam.xfam.org>). Hmsearch, which is part of the HMMER package (v3.1b2) (35), with the Gathering Cut-Off threshold (-cut_ga) was used to identify occurrences of these domains in the predicted proteomes of all 109 fungal species considered here. To filter out short and incomplete fragments, we applied a cutoff calculated by 50% length coverage of the query sequence (domain) or the hit. Subsequently, the identified domain sequences were aligned using MAFFT v7.271 (36). TrimAl v1.2 (-gt 0.1) was used to remove positions in the alignments with gaps in more than 90% of the sequences (37). Maximum-likelihood phylogenetic trees were constructed using IQ-TREE v1.6.10 (38,39) using ultrafast bootstrap and Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) (40) with 1,000 pseudo-replicates, and standard model selection to automatically determine the best-fit model. We predicted additional domains in the analyzed proteins using Hmsearch with all Pfam domain profiles that are present in the Pfam database (v28.0) and visualized them along the phylogenetic trees using iTOL (41).

Species tree construction

To construct a species phylogeny of the here analyzed fungi, we used BUSCO v4.0.1 (Benchmarking Universal Single-Copy Ortholog assessment tool) (42) to identify the occurrence of 758 fungal single-copy genes (Supplementary Table 2) within the 109 predicted proteomes (Supplementary Table 1). The identified homologs of single-copy BUSCO genes were subsequently concatenated into a single super-alignment. Then, the maximum-likelihood species phylogeny was estimated using IQ-TREE v1.6.10 (38,39) by employing partitioned model selection, which allows selecting a substitution model for each BUSCO gene. Branch supports were obtained using ultrafast bootstrap and SH-aLRT (40) with 1,000 pseudo-replicates separately.

Chromatin modifier subunits homology search and gene tree construction

We obtained previously studied chromatin modifier complexes from literature, and the associated protein sequences were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov>) (Supplementary Table 3). For each subunit of the chromatin modifier complexes, homologs in the analyzed 109 predicted proteomes were detected based on sequence similarity searches. First, BLASTP (v2.2.31+) (43) searches were performed, and an e-value threshold of 10^{-2} was applied. Multiple BLAST matches (High-Scoring Pairs: HSPs) of the same query were concatenated if they were aligned with different non-overlapping regions. Only the non-redundant region was retained when HSPs of the same query were overlapping. For these two conditions, the highest e-value of the different matches was used to represent the merged match. Only matches with the e-value lower than 10^{-5} and covering more than 50% of query or subject were retained for the following analyses. If we observed less than ten hits for the BLASTP search after filtering in Aspergilli, we used PSI-BLAST (Position-Specific Initiated BLAST v2.2.31+) (43) with five iterations (-num_iterations 5) using the best Aspergilli hit from the BLASTP as the query; if no Aspergilli match was retained, we used the same query as for the BLASTP searches. The same filter criterion as for the BLASTP selection was applied to

the PSI-BLAST hits, and the results of these two homology search strategies were subsequently jointed.

To account for apparent protein absences due to erroneous gene annotation, we tested these potential losses by using TBLASTN (BLAST v2.2.31+) (43) to query protein sequences against the genome assemblies. Matches with e-value lower than 10^{-5} together with 4,500 nucleotides upstream and downstream were extracted, and Exonerate v2.2.0 (44) was used to predict the gene structure (options: protein2genome, extensive search, showtargetgff, showcigar, and showquerygff options were set as true, and “>%ti(%tcb - %tce)\n%tcs\n” option was used to output the complete aligned coding sequence). Transeq v6.6.0.0 (45) was used to translate the coding sequence into protein sequences.

Together with the protein sequences identified from BLASTP and PSI-BLAST, all protein sequences from each subunit were retrieved to build a phylogenetic tree (MAFFT, TrimAl, and IQ-TREE settings were the same as for the tree reconstruction using the catalytic domains above).

For each sequence included in the phylogenetic trees, the domain organization was identified and visualized as described above. When the domain organization of a protein differed from its orthologues, the corresponding gene model was manually checked in the MycoCosm genome browser (1) and corrected. After manual curation, the presence and absence patterns of 86 subunits from 16 different chromatin modifier complexes were visualized using the Seaborn library in Python (46,47).

Histone enrichment and mass spectrometry

Spores of 4- or 5-day old sporulating *A. nidulans* (AnWT pabaA1 veA1, kindly provided by Prof. Joseph Strauss) growing on ME (Malt Extract) agar medium were harvested with 2 mL

sterile water and cultured in 50 mL ME liquid medium at 28 °C, 200 rpm for 16 hours. The sample was centrifuged at 8,000 g for 10 minutes to discard supernatant, and germinating spores were resuspended in 25 mL 0.8 M NaCl and centrifuged again. Ten mL protoplasting solution (20 mg/mL *Trichoderma* lysing enzyme and 5 mg/mL driselease in 0.8 M NaCl) were used to resuspend spores, which were then incubated at 30 °C, 100 rpm for 2-3 hours. Protoplasts were retrieved by filtering through sterile Miracloth and were concentrated by centrifugation at 3,000 rpm, 4 °C for 5 minutes. Histone proteins were extracted according to an adapted histone protein extraction protocol (48). Briefly, 10⁸ protoplasts were resuspended in 1 mL Nuclei Isolation Buffer (1×PBS, 0.5 mM PMSF, 5 µM Leupeptin, 5 µM Aprotinin, and 5mM Na-butyrate) with 1% Triton X-100, followed by vigorous pipetting through a 200 µL pipette several times. Then, centrifuged at 2,300 g for 15 minutes at 4 °C, discarded the supernatant, and resuspended the nuclear pellet in 200 µL Nuclei Isolation Buffer with 0.1% SDS and 250 U Pierce™ Universal Nuclease and incubated at 37 °C for 30 minutes to digest nucleic acids. The protein concentration was measured using the Bradford Protein Assay (49).

The samples were separated on a 17% SDS-PAGE gel and histone bands were excised, chemically acylated with propionic anhydride and in-gel digested with trypsin, followed by peptide N-terminal derivatization with phenyl isocyanate (PIC) (50). The samples were then desalted on handmade StageTips columns (51). Peptide mixtures were separated by reversed-phase chromatography on an EASY-Spray column (Thermo Fisher Scientific), 25-cm long (inner diameter 75 µm, PepMap C18, 2 µm particles), which was connected online to a Q Exactive Plus instrument (Thermo Fisher Scientific) through an EASY-Spray™ Ion Source (Thermo Fisher Scientific), as described (50).

The acquired RAW data were analyzed using the integrated MaxQuant software v.1.6.10. The Uniprot UP000000560 database was used for identification of *A. nidulans* histone peptides. Enzyme specificity was set to Arg-C. The estimated false discovery rate

(FDR) was set at a maximum of 1%. The mass tolerance was set to 6 ppm for precursor and fragment ions. Two missed cleavages were allowed, and the minimum peptide length was set to 4 amino acids. Min. score for modified peptides and min. delta score for modified peptides were set to 1. Variable modifications included lysine propionylation, monomethylation + propionylation, dimethylation, trimethylation and acetylation. N-terminal PIC labeling was set as a fixed modification (50). To reduce the search time and the rate of false positives, with increasing the number of variable modifications included in the database search, the raw data were analyzed through multiple parallel MaxQuant jobs (52), setting different combinations of variable modifications. Peptides identified by MaxQuant with Andromeda score higher than 50 and localization probability score higher than 0.75 were quantitated, either manually or by using a version of the EpiProfile 2.0 software (53) adapted to the analysis of histones from *A. nidulans*. Identifications and retention times were used to guide the manual quantification of each modified peptide using QualBrowser version 2.0.7 (Thermo Fisher Scientific). Site assignment was evaluated from MS2 spectra using QualBrowser and MaxQuant Viewer. Extracted ion chromatograms (XICs) were constructed for each doubly charged precursor, based on its m/z value, using a mass tolerance of 10 ppm. For each histone modified peptide, the % relative abundance (%RA) was estimated by dividing the area under the curve (AUC) of each modified peptide for the sum of the areas corresponding to all the observed forms of that peptide (54). The AUC values are reported (Supplementary Table 4) and visualized using GraphPad Prism.

Results and discussion

DNA methylation, histone methylation, and histone acetylation are among the most studied chromatin PTMs in animals, plants, and fungi (55–57). Protein complexes involved in writing, reading, and erasing these modifications have been reported in very few model fungal species to date (16,22), and the evolution and conservation of these complexes in diverse fungi has not been systematically explored. To start uncovering the occurrence and evolutionary histories of chromatin modifiers involved in these major modifications in the fungal genus *Aspergillus*, we first focused on five conserved catalytic domains reported to be involved in histone PTMs: DNA methyltransferase domain for DNMTs, SET domain for histone methyltransferases (HMTs), acetyltransferase domain and MOZ/SAS FAMILY domain for histone acetyltransferases (HATs), and histone deacetylase domain for histone deacetylases (HDACs). We identified homologs harboring these conserved domains in 109 predicted proteomes; the completeness of the predicted proteomes was generally high as 97 had more than 97% complete BUSCO predicted proteins (Supplementary Table 1). The 94 *Aspergilli* species cover 18 recognized sections: *Polypaecilum*, *Cremeri*, *Restricti*, *Aspergillus*, *Cervini*, *Clavati*, *Fumigati*, *Candidi*, *Circumdati*, *Janorum*, *Terrei*, *Flavi*, *Ochraceorosei*, *Bispori*, *Cavernicolarum*, *Usti*, *Nidulantes*, and *Nigri* (58–64), and 15 other species spanning the Dikarya were used as outgroups to the *Aspergillus* genus (Figure S1). We used the identified homologs for each catalytic domain to determine their evolutionary history using maximum-likelihood phylogenies (Figure 1-4).

Aspergilli harbor conserved DNA methyltransferases for genome defense and methylation maintenance

The conserved DNA methyltransferase domain is the only domain known to catalyze the addition of methyl groups to DNA specifically at cytosines (5mC) (26). Based on previous domain architecture analyses, phylogeny reconstructions, and functional experiments, five distinct protein families containing a DNA methyltransferase domain have been recognized in fungi: DNMT1, DNMT2, and DNMT5 are commonly found in eukaryotes, while Dim-2 and RID

Figure 1. Conservation of DNA methyltransferase families in the *Aspergillus* genus. The maximum-likelihood phylogeny of DNA methyltransferase domains (PF00145) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in DNA methyltransferase domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

are considered to be fungal specific (26). Among these five families, DNMT2 functions as a tRNA methyltransferase rather than DNA methyltransferase (65), and it only occurs in our study in two Basidiomycota and one Ascomycota (*Schizosaccharomyces pombe*), which is in line with previous observation suggesting that it has been lost in most Ascomycota (26,66).

Three (*DNMT1*, *DNMT5*, and *Dim-2*) of the four remaining families catalyze the maintenance of DNA methylation pattern from parental DNA strand onto synthesized daughter DNA strands (67,68). Based on our species selection and phylogenetic analyses, we here found *DNMT1* and *Dim-2* to be absent in all Aspergilli and only present in four or five out of 15 outgroup species, respectively, while *DNMT5* is found in most Aspergilli and outgroup species (Figure 1). Only four outgroup species harboring *DNMT1*, the two Basidiomycota species, *Botrytis cinerea*, and *Plectania melastoma*, which corroborates previous findings (26,69); and *Dim-2* was identified in only five Ascomycota outgroup species. Recently, it was found that *Dim-2* is inactivated in some *Zymoseptoria tritici* isolates only, which uncovered that *Dim-2* plays a role in genome defense and genome evolution (70). In contrast to the *DNMT1* and *Dim-2* losses in Aspergilli, *DNMT5* is identified in 88 out of 94 Aspergilli and 10 out of 15 outgroup species (Figure 1), with a conserved domain architecture consisting of a DNA methyltransferase domain, an SNF2 family domain (PF00176), and a Helicase conserved C-terminal domain (PF00271), in which the latter two domains are involved in transcription regulation (71). The five outgroup species without *DNMT5* are *S. cerevisiae*, *S. pombe*, *Neurospora crassa*, *Fusarium graminearum*, and *Cladonia grayi*, among which the two yeast species also lack the other two DNA methylation maintaining enzymes, and thus accordingly are considered to be incapable of DNA methylation (72–74). Five out of the six *Aspergillus* species that lost *DNMT5* belong to a single clade within the *Nidulantes* section (*A. undulatus*, *A. olivicola*, *A. filifera*, *A. venezuelensis*, and *A. stella-maris*), and the sixth one is *A. robustus* which is assigned to the *Circumdati* section (Supplementary Figure 1). In the Basidiomycota fungus *Cryptococcus neoformans*, *DNMT5* maintains DNA methylation using the same cofactors as *DNMT1* (73,75). Thus, it is most likely that *DNMT5* is similarly involved in DNA methylation maintenance in Aspergilli as in *C. neoformans*. Moreover, DNA methylation maintenance does not appear to be required for the survival of fungi as shown with the absence of all three *DNMT1*, *DNMT5*, and *Dim-2* not only within the *Nidulantes* section but also in *S. cerevisiae* and *S. pombe* (Figure 1). However, the absence of these genes could have an impact on the regulation of certain genes like secondary metabolite biosynthetic gene

clusters and it would be interesting to compare species within the *Nidulantes* section in more detail to identify how such loss may have affected gene expression and genome organization.

RID is a more recently acquired single domain DNMT that specifically occurs in fungi and is involved in a fungal-specific genome defense mechanism called RIP (Repeat-Induced Point mutation) (26). RID is required for *de novo* DNA methylation of transposable elements (TEs) in both CG and non-CG context relating to RIP during the sexual stage in *N. crassa*, but its *de novo* DNA methylation capacity during the vegetative growth has not been reported (76). Here, we identified RID in all analyzed Aspergilli species as well as in ten outgroup species. A RID-like homolog knock-out mutant in the model filamentous Ascomycota *Podospora anserina* shows blocked sexual development, as gametes formation and fertilization can happen but dikaryotic cells cannot develop normally (77). Interestingly, the knock-out mutant of *DmtA* (homolog of *RID* in *A. flavus*) shows altered conidiation, sclerotial production, aflatoxin biosynthesis, and virulence (32), suggesting that RID might have additional functions next to genome defense. Additionally, fungi lost the *de novo* DNA methyltransferase *DNMT3* around 50–150 million years ago (75,78), and only *DNMT5* and *RID* were consistently found in Aspergilli (Figure 1). Thus, we hypothesize that Aspergilli can only maintain DNA methylation *via* DNMT5 and may generate *de novo* DNA methylation *via* RID during the transient sexual stages, but do not spur genome-wide *de novo* DNA methylation. This hypothesis seems to be corroborated as bisulfite sequencing in *A. flavus* has uncovered very little genome-wide DNA methylation (79). However, systematic analyses of DNA methylation patterns and experimental evidence for the roles of DNMTs in Aspergilli are still largely missing.

Aspergilli lack SET7, but SET-domain proteins are abundant and histone methyltransferases are conserved

Histone methylation is typically catalyzed by SET domain-containing proteins (80). We identified a total of 1,548 SET domain-containing proteins and classified them into 15 groups based on their domain architecture and phylogenetic support that contain species both from

Figure 2. Conservation of histone methyltransferase families in the *Aspergillus* genus. The maximum-likelihood phylogeny of SET domains (PF00856) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles and leaves with similar protein architecture with unknown functions covering both *Aspergilli* and outgroup species are collapsed and shown as grey triangles. Additional conserved domains in SET domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Aspergilli as well as from the outgroup. Among these groups, 107, 111, 109, 102, and 108 sequences represent homologs of the known histone methyltransferases (HMTs) SET9, Dim-5, SET1, Ash1, and SET2, respectively (Figure 2). The remaining homologs were divided into nine groups that we broadly classified as SET domain-containing proteins. Among them, eight groups contain a SET domain but lack experimentally determined functions, and 117 sequences could be grouped as Ribosomal lysine N-methyltransferases (Figure 2). Enzymes in this group contain a SET domain together with a Rubisco LSMT substrate-binding domain (PF09273). While the homolog in plant transfers a methyl group to the large subunit of the Rubisco holoenzyme complex (81,82), the homolog in *S. cerevisiae* has been characterized as a ribosomal lysine N-methyltransferase (83).

The well-described HMTs SET1, Dim-5, SET2, Ash1, and SET9 all contain distinct accessory domains and catalyze methylation on lysine residues at different positions of H3 or H4 tails (22). SET1 is composed of the SET domain, the N-SET domain (PF11764), as well as the SET_assoc domain (PF11767), and is responsible for H3K4me1/2/3 in *S. cerevisiae* (84). SET1 occurs in a single copy in all species (Figure 2), indicating that this enzyme and its catalyzed PTM play fundamental roles in fungal biology. Dim-5 contains a SET domain and Pre-SET domain (PF05033) and acts as an H3K9 methyltransferase inducing heterochromatin formation in *N. crassa* (85). Dim-5 was found in all species included in this study except *S. cerevisiae* and *A. calidoustus*, among which *S. cerevisiae* is reported to lack H3K9me (85,86). How such a loss has impacted gene expression in *A. calidoustus* in comparison to its closest relative *A. carlsbadensis* deserves further investigation. Both SET2 and Ash1 catalyze H3K36me in *N. crassa* and share the SET domain and the AWS domain (PF17907) (Figure 2), yet SET2 harbor one additional SRI domain (PF08236) that interacts with RNA polymerase II and links H3K36me with transcript elongation (87–89). In *N. crassa*, Ash1 was shown to contribute to H3K36me at sub-telomeric regions, while SET2 acts primarily at euchromatic regions (90,91). Considering the common origin of SET2 and Ash1 as well as their similar activities, an ancient duplication followed by the loss or gain of the SRI domain is likely to have

contributed to their functional diversification (Figure 2). SET2 was found in all outgroup species, while *Ash1* was lost in both Basidiomycota, both yeasts, *B. cinerea*, and *F. graminearum*. Both enzymes are conserved in all *Aspergillus* species except in *A. carbonarius* from the *Nigri* section, suggesting that either other SET domain-containing proteins exhibit functional redundancy or that H3K36 methylation is not fully required in this species. In *Fusarium fujikuroi*, deletion of *SET2* and *ASH1* strongly affected vegetative growth and conidiation (91). SET9 homologs harbor only a SET domain, and were shown to catalyze H4K20me in *S. pombe* (92). This group is well conserved in most of the species except in *S. cerevisiae* (93) and in the two closely related Aspergilli, *A. novoparasiticus* and *A. arachidicola* that both belong to section *Flavi* (Figure 2 and Figure S1).

SET7 is the catalytic subunit of PRC2, and is highly conserved in mammals, plants, and *N. crassa*, in which it has been well-characterized to catalyze methylation on H3K27, a PTM that is deposited at facultative heterochromatin (21). Even though PRC2 is considered to be a key chromatin modifier complex, its catalytic subunit *SET7* was lost in many fungal species (22). *SET7* has previously been reported to be absent in *A. nidulans* and *A. fumigatus* (20–22), and we here demonstrate that *SET7* is absent from all analyzed Aspergilli as well as from *Penicillium* (Figure 2), suggesting that this enzyme was most likely lost at or before the divergence of the last common ancestor of these two fungal genera. Thus, in line with previous reports (22,30,94,95), our results indicate that methylation on H3K4, H3K9, H3K36, and H4K20 are likely catalyzed by conserved enzymes in all Aspergilli, while H3K27 methylation is absent. Considering that loss of function of these enzymes are usually reported to trigger deleterious phenotypes (32,77), investigating how some species are able to cope with such losses will likely provide new insights on the histone code and its conservation in fungi. Importantly, the occurrence of a high number of uncharacterized SET domain-containing proteins opens the possibility that these contribute to not yet characterized histone modifications. Especially, four uncharacterized SET domain-containing groups with 144, 195,

113, and 113 members are conserved in *Aspergilli* and thus likely play important roles in different cellular processes.

Histone acetyltransferases are conserved in *Aspergilli*

Histone acetyltransferases (HATs) have been classified into five families: MYST ('MOZ, Ybf2/Sas3, Sas2, and Tip60)-related HATs, Gcn5-related acetyltransferases (GNATs), p300/CBP HATs, general transcription factor HATs, and nuclear hormone-related HATs (96–98). We here only focused on the first two HAT families because these are known to be involved in the regulation of histone proteins and appear largely conserved in fungi (96).

The MOZ/SAS FAMILY domain is the catalytic domain of MYST-related HATs (99). As previously reported (100), the three characterized MYST-related HATs Sas3, Sas2, and Esa1 belong to distinct monophyletic clades, and these paralogues are nearly fully conserved in the here analyzed fungi (Figure 3). MOZ/SAS FAMILY domain and zf-MYST domain (PF17772) are present in Sas3, which can on its own acetylate histone H3 and H4, and H2A weakly *in vitro* (101). Within the NuA3 (Nucleosomal Acetyltransferase of histone H3) complex, Sas3 histone acetylation activity is *in vitro* restricted on H3K14 and to a less extent on H3K23, while in *S. cerevisiae*, only H3K14ac can be detected (102). Sas3 occurs in most of the species but is lost in *S. pombe*, as previously observed (103,104), and is duplicated in *A. carbonarius* (Figure 3). In addition to the two domains found in Sas3, Sas2 and Esa1 carry a PHD_4 domain (PF16866) or Tudor-knot domain (PF11717), respectively (Figure 3). Esa1 is present as a single copy in all species included in this study, and its functions is known to contribute to the acetylation of K5, K8, K12, K16 on histone H4, as well as on H2AZK14 in *S. cerevisiae* (105–108). Sas2 is similarly found in most of the species except for two Basidiomycota as well as for *A. carbonarius*. Sas2 is involved in chromatin-mediated gene regulation and H4K16 acetylation in both *S. cerevisiae* and *Candida albicans* (105,109,110). Although Esa1 and Sas2 catalyze H4K16ac, it was shown in *C. albicans* that they are differentially recruited at different stages of development (105). Because both Esa1 and Sas2 are well-conserved in

Figure 3. Conservation of MOZ/SAS FAMILY domain-containing histone acetyltransferases in the *Aspergillus* genus. The maximum-likelihood phylogeny of MOZ/SAS FAMILY domains (PF01853) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in MOZ/SAS FAMILY domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

the *Aspergillus* genus, we hypothesize that these two HATs may also have distinct roles during development.

GNATs have an acetyltransferase domain as its catalytic core, and 2,604 acetyltransferase domain-containing proteins were found in our search (Supplementary Figure 2). The two well-characterized GNATs, *Gcn5* and *Elp3*, occur in a single copy in every species included in our study (Supplementary Figure 2). *Gcn5* harbors a Bromodomain (PF00439) in addition to the Acetyltransferase domain and is a crucial subunit of distinct histone acetyltransferase complexes such as the HAT-A2 complex, the SAGA complex, and the ADA complex in *S. cerevisiae* (96,98). All these complexes contain two coactivator proteins, Ada2 and Ada3, that directly bind to *Gcn5* (111) and the recruitment of different other subunits determines the location where they catalyze acetylation (96). *Elp3* is characterized by the Acetyltransferase domain, Radical SAM (PF04055), and Radical SAM C-terminal domain (PF16199), and has been shown to catalyze H3K14ac, H4K8ac, and H3K56ac in *S. cerevisiae* (112–114). *Elp3* also exhibits a dual enzymatic activity as it catalyzes demethylation in mouse zygotes (115), which is due to the Radical SAM domain that catalyzes the demethylation of methylated lysyl residues on histones (116).

Next to *Gcn5* and *Elp3*, we have identified 2,386 additional acetyltransferase domain-containing proteins. Because their functions remain unknown due to the lack of homology to sequences with characterized functions, they are valuable candidates that may acetylate histones or other proteins. Interestingly, it has been shown that *Gcn5* also exhibits succinyltransferase activity in humans (117) and can also function as histone crotonyltransferase to regulate gene transcription *in vitro* (118). It is expected that other less common histone modifications, *i.e.* histone lysine butyrylation, propionylation, or malonylation, may also be catalyzed by histone acetyltransferases because these histone modifications all use acyl-Coenzyme A to complete the reactions (119–121). Further research is needed to determine whether these less common histone modifications are catalyzed by the known histone acetyltransferases *Gcn5* and *Elp3*, or as of yet unstudied HATs. In conclusion, all five characterized fungal HATs described above are highly conserved in the *Aspergillus* genus, showing their importance in fungal biology. Many other acetyl-transferase domain-containing

proteins are encoded in *Aspergillus* genomes and may be responsible for histone acetylation or additional less common histone modifications like butyrylation, succinylation, and malonylation.

Duplication of *RpdA* histone deacetylase gene in the *Flavi* and *Janorum* sections

Five families of histone deacetylases (HDACs) can be grouped into two classes based on sequence similarity: class I (HOS1, HOS2, and Rpd3/RpdA) and class II (HdaA and HOS3) (122). They all harbor the Histone deacetylase catalytic domain, and our phylogenetic tree corroborates the HDAC classification into two distinct classes (Figure 4). *HdaA* is found as a single copy in each species included in this study (Figure 4). It harbors the Histone deacetylase domain and C-terminal Arb2 domain (PF09757) that serves as an anchor to target centromeric heterochromatin region (123,124). *HdaA* deacetylates lysine in histones H3 and H2B, but not H4 nor H2A in *S. cerevisiae* (125,126), and the $\Delta hdaA$ knock-out strains show increased or reduced production of different secondary metabolites in *A. nidulans* (127), *A. niger* (127), and *A. fumigatus* (128). By contrast, *HOS3* is found in only 80 *Aspergillus* species, as it is lost in nine species (*A. nomiae*, *A. pseudocaelatus*, *A. tamarii*, *A. pseudotamarii*, *A. caelatus*, *A. sergii*, *A. novoparasiticus*, *A. minisclerotigenes*, *A. arachidicola*) in the *Flavi* section, three species (*A. desertorum*, *A. stercorarius*, and *A. multicolor*) in the *Nidulantes* section, and two species (*A. heteromorphus* and *A. carbonarius*) in the *Nigri* section. *HOS3* is well conserved in all outgroup species except *S. pombe* (Figure 4) (129). *HOS3* has been shown to catalyze deacetylation with distinct specificity for histone H2AK7, H2BK11, H3K14 and H3K23, as well as H4K5 and H4K8 in *S. cerevisiae* cell extracts (130). Notably, *HOS1* was only detected in three outgroup species, *S. cerevisiae* and both Basidiomycota (Figure 4). The specificity of *Hos1* in catalyzing deacetylation on histone proteins is not known yet (131), but it can catalyze the deacetylation of the Smc3 subunit of Cohesin and influence the regulation of chromosome segregation during mitosis in *S. cerevisiae* (132). *HOS2* is found in 15 outgroup species and most of *Aspergilli* except *A. oryzae* (Figure 4). It is known to deacetylate lysine residues in H3 and H4 histone tails (133). The conservation of *HOS2* in

Figure 4. Conservation of histone deacetylase domain-containing proteins in the *Aspergillus* genus. The maximum-likelihood phylogeny of Histone deacetylase domains (PF00850) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (37,38). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in histone deacetylase domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Aspergilli suggests that it plays a key role, which is supported by the deletion of *HOS2* in *A. niger* where the deletion strains displays severe reduction in growth, sporulation, SM biosynthesis, and stress resistance (127).

Within class I HDACs, the Rpd3/RpdA clade is divided into two sub-clades (Figure 4). One is nearly fully conserved (except in *A. oryzae*) and comprises the well-characterized Rpd3/RpdA HDACs, while the other one only comprises sequences from the *Flavi* and *Janorum* sections (Figure 4). Our phylogenetic reconstruction suggests a duplication event that has occurred in the ancestor of all Aspergilli giving rise to a paralog that was retained in two sections only and could explain the *A. oryzae* loss in the conserved sub-clade. Rpd3 is a crucial deacetylase in eukaryotes, as it is required for deacetylation in most locations on core histone proteins, except H4K16 *in vitro* (134) that is catalyzed by HOS2 (133). *S. cerevisiae* lacking *Rpd3* is sensitive to high osmolarity and shows compromised expression of osmotic stress genes (135). Deletion of *RpdA* in *A. nidulans* leads to the reduction of its growth and sporulation (136), while only heterokaryon disruptants could be obtained in *A. oryzae* (137), indicating its crucial role in the fungal genus *Aspergillus*. In summary, class I and class II deacetylases are well conserved in Aspergilli, and duplication events like the one observed for *RpdA* may confer new activities. While only one paralog was retained in *A. oryzae*, the presence of two paralogs in sections *Flavi* and *Janorum* may provide functional redundancy or diversification, which remains to be determined.

Presence-absence patterns of subunits that compose histone modifier complexes

Histone PTMs are catalyzed by the conserved domains found in DNMTs, HMTs, HATs, and HDACs (Figure 1-4), but these chromatin modifying enzymes are often not acting alone. Instead, they are part of protein complexes that comprise accessory subunits, which are involved in the recognition and induction of specific histone modifications at the correct location (18,19). To further investigate the presence-absence patterns (PAPs) of known chromatin modifier complexes in Aspergilli, we focused on 16 chromatin modifier complexes

with 86 subunits previously identified and studied in *S. cerevisiae* or *N. crassa* (Supplementary Table 3). Apart from the Velvet complex, whenever a catalytic subunit is absent in one species, there seems to be a trend for the absence of additional accessory subunits, which is particularly observed in outgroup species (Figure 5). Only a few PAPs are observed, and most accessory subunits are either fully conserved or absent in Aspergilli. Nine subunits appear specific to *S. cerevisiae* as they are not found in any other species and are likely to play a specific role in this species. Seven subunits are absent in Aspergilli and show PAP in the outgroup species.

Five histone methyltransferase complexes, COMPASS, DCDC (Dim-5/-7/-9/CUL4/DDB1 Complex), RNAP2, PRC2, and Dot1 (disruptor of telomeric silencing 1), as well as the putative VelB/VeA/LaeA methyltransferase complex (Velvet), were included in our analyses (Figure 5). Five COMPASS subunits (*SET1*, *CclA*, *Swd1*, *Swd2*, and *Swd3*) are fully conserved, suggesting an important function for H3K4 methylation. Consistent with a PAP in most outgroup species and the loss in Aspergilli, Spp1p is not essential for H3K4 methylation (138). As *Sdc1* shows a PAP in Aspergilli and outgroup species, it is likely that this subunit also does not play a key role in H3K4 methylation. In the DCDC complex, while Dim-7 and Dim-9 seem to be dispensable for H3K9 methylation, *Dim-5*, *Cul4*, and *Dim-8* are highly conserved, and we thus expect these enzymes to play key roles in the activity of this complex (Figure 5 and Table S3). However, it was shown that Dim-7 and Dim-9 are responsible for heterochromatin recognition, which leads to subsequent recruitment of the complete DCDC to induce H3K9 methylation in *N. crassa* (139). Thus, it is very likely that other accessory subunits are involved in recruiting DCDC at specific chromatin locations in Aspergilli. Furthermore, the unique absence of *Dim-8* in *A. indicus* may suggest altered H3K9 methylation in this species. Consistent with the absence of *SET7* (Figure 2), all components of the PRC2 complex are lost in Aspergilli (Figure 5). *PRC2* is found in different filamentous fungi such as *N. crassa* (21), *Z. tritici*, *B. cinerea*, and *F. graminearum* (22). The apparent loss

Figure 5. Presence-absence pattern of subunits that constitute characterized histone modification complexes. The top row with distinct color boxes indicates outgroup species and Aspergilli sections, the order of the species from right to left is the same as in the species tree from top to bottom (Supplementary Figure 1) without *Fusarium oxysporum* and *Aspergillus kawachii* since they were excluded in the later analysis. Each row represents a subunit in the histone modification complexes, the order agrees with the summary list (Supplementary Table 3). Each box in the matrix indicates the presence (dark color) or absence (light color) of a subunit (Supplementary Table 3). Each row shows the presence/absence of a specific complex subunit. For each complex, the first row represents the catalytic subunit.

of *SET7* in *B. cinerea* and *Z. tritici* in our study (Figure 5) is due to their length that was too short to pass the inclusion threshold compared to the *N. crassa SET7* query.

We included eight histone acetyltransferase complexes that differ by their distinct catalytic subunits (Figure 5). The HAT-A2, ADA, SAGA, and Elongator complexes use an acetyltransferase domain-containing protein as catalytic subunit (34,140). The SAS (Something About Silencing), NuA4 (Nucleosomal Acetyltransferase of histone H4), and NuA3 complexes use a MOZ/SAS FAMILY domain-containing protein (141). The NuB4 (Nuclear Hat1p-containing type B histone acetyltransferase) complex uses the histone acetyltransferase HAT1 domain-containing protein (142). Gcn5, Ada2, and Ada3 are fully conserved in all fungi and constitute the core enzymatic module of the HAT-A2, ADA, and SAGA complexes. This catalytic module contributes to catalyzing histone acetylation, and other subunits in these complexes guide them to specific locations and link them to the transcriptional machinery (143). Besides the catalytic module, the ADA complex also harbors the Ahc1 subunit that is found in *S. cerevisiae* only (140,144). Twelve additional subunits of the SAGA complex are well conserved in Aspergilli and outgroup species (Figure 5). The SAGA complex also contains a deubiquitylase (DUB) module, for which the composition differs between Aspergilli and other species. *Sgf11* is unique to *S. cerevisiae* and *Sus1* was likely lost in the ancestor of Pezizomycotina because we can only identify *Sus1* orthologs in both Basidiomycota and the yeasts (Figure 5). The Elongator complex appears more variable as four subunits are mostly conserved, but *Elp6* is specific to *S. cerevisiae*, and *Hpa2* and *Hpa3* show PAPs in both Aspergilli and outgroup species.

The NuA4, SAS, and NuB4 complexes are well-conserved in Aspergilli as well as in the outgroup species (Figure 5). These complexes appear slightly different in *S. cerevisiae* with the presence of specific subunits (*Eaf5* in the NuA4 complex; *Sas4* and *Sas5* in the SAS complex). Both SAS and NuA4 complexes can catalyze H4K16ac with opposite effects during cell growth and morphogenesis in *C. albicans* (105). It would be interesting to assess whether these two complexes exhibit similar activities in Aspergilli. For the NuA3 complex, two accessory subunits are well conserved among Aspergilli while two others, *Pdp3*, and *Nto1*, show different PAPs, and *Yng1* is specific to *S. cerevisiae*. *Pdp3* and *Nto1* seem to be lost in

a complementary manner, suggesting a possible functional redundancy (Figure 5). Finally, two histone deacetylase complexes, Sin3-Rpd3 and HCHC, are fully conserved (Figure 5). The Sin3-Rpd3 complex is slightly different in *S. cerevisiae* with the specific of the *Ume1* subunit.

Overall, our analysis showed that the core catalytic modules of complexes involved in histone PTMs are mostly conserved in all fungi, including Aspergilli, but variability in accessory subunits could contribute to differences in the localization of histone PTMs along the genome.

Unbiased mass spectrometry captures the diverse histone modifications in *A. nidulans* and verifies the absence of H3K27 methylation

Our genomic and phylogenetic analyses have identified the presence of multiple chromatin modifiers in Aspergilli, suggesting that these species have the ability to establish a wide variety of histone modifications. To assess the presence of histone PTMs predicted in our evolutionary analyses (Fig 1-5), we extracted histone proteins from *A. nidulans* and performed unbiased mass spectrometry analyses. H3 and H4 peptides with PTMs corresponding to H3K4me, H3K9me, H3K36me, and H3K79me were detected (Figure 6), which we anticipate being catalyzed by SET1, Dim-5, Ash1 and SET2, and Dot1, respectively. Interestingly, H4K20me, which is catalyzed by SET9 in *S. pombe* (92), was not detected (Figure 6). However, *SET9* is well conserved in Aspergilli (Figure 2), and thus it is possible that Aspergilli lost subunits that are needed to induce this specific PTM or that H4K20me is not detectable under the here tested conditions. Additionally, acetylation of H3K9, H3K14, H3K18, H4K5, H4K8, H4K12, and H4K16 were detected (Figure 6). Acetylation on H3K23 and H3K56 were not detected in this study even though the core catalytic enzymes of SAGA, NuA3, and Elongator complexes involved in these modifications are conserved (Figure 5). These modifications either do not occur in Aspergilli, maybe due to the absence of the necessary accessory subunits or occur below our detection threshold. As expected, we did not observe any peptide that would indicate the presence of H3K27me in *A. nidulans* (Figure 6A;

Supplementary Table 4), thereby corroborating our computational analysis that the PRC2 complex is absent in *A. nidulans* as well as in all other Aspergilli. Thus, the mass spectrometry results verify the occurrence of histone modifications predicted to be present based on our comparative analyses in the fungal genus *Aspergillus*. Moreover, the differences between the relative abundance of distinct histone modifications provide interesting information to investigate these abundances and identify specific conditions under which these modifications increase or decrease in abundance.

Figure 6. Schematic representation of the histone code in Aspergilli. (A) Relative abundance of post-translational modifications detected on peptides of histone H3 and H4 from *Aspergillus nidulans* using unbiased mass spectrometry. The term 'combinations' refers to multiple modifications present on the same peptide (Supplementary Table 4). (B) Histone modifications are predicted by genomic and phylogenetic analyses, and/or proved by mass spectrometry in *A. nidulans*. Histone modifications verified by mass spectrometry are labeled as solid dots, while the ones that are only predicted by phylogenetic analysis are labeled as solid dots with a red outline.

Conclusion

Changes in chromatin organization play an important role in regulating the expression of genes (129). The fungal genus *Aspergillus* is well known for its ability to thrive in a broad range of different environmental niches and the huge potential of secondary metabolite biosynthesis (16,145). However, a detailed and complete overview of occurrence and evolution of chromatin modifiers in Aspergilli is a prerequisite to be able to better understand how these fungi regulate gene expression. In the future, more research is needed to uncover how the absence of *de novo* DNA methylation impacts genome evolution in Aspergilli. We demonstrate that PRC2 is lost throughout Aspergilli, and the absence of PRC2 is responsible for the lack of H3K27me. Although histone acetyltransferases and deacetylases are overall conserved, our study also revealed specific duplication and loss events, which require further investigation to understand their impact on chromatin and gene regulation in certain *Aspergillus* species. Moreover, other less common histone modifications like histone lysine succinylation, crotonylation, butyrylation, propionylation, and malonylation, are all likely catalyzed by HATs and HDACs as they all use Coenzyme A (CoA) to complete the reactions (119–121), but may require specific yet to be discovered subunits. In conclusion, our study provides a detailed overview of the evolutionary routes of chromatin modifier complex in the fungal genus *Aspergillus*, and therefore generates the necessary framework to identify targets for functional studies to understand how chromatin regulates gene expression in Aspergilli.

Acknowledgment

The sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community. Especially, we thank the Scott Baker, Mikael Andersen, and Joseph Spatafora for their permission to use the sequence data. We also like to thank Jos Houbraken for validating the *Aspergillus* species tree, Joseph Strauss for providing *A. nidulans* strain, and Alessandro Vai for supporting the mass spectrometry data analysis of histone PTMs.

Funding information

Xin Zhang is funded by the Chinese Scholarship Council (CSC) (201907720028). This work has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 programme of the European Union.

Authors and contribution

X.Z. formal analysis, investigation, methodology, visualization, writing – original draft; R.N. formal analysis, investigation, methodology, visualization; T.B. funding acquisition, supervision, resources; J.C. Conceptualization, funding acquisition, methodology, project administrations, supervision, visualization, writing original draft; M.F.S. Conceptualization, funding acquisition, methodology, project administrations, supervision, visualization, writing original draft.

References

1. Grigoriev I V, Nikitin R, Haridas S, Kuo A, Ohm R, Otilar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 2014;42(D1):D699–704.
2. Labrador M, Corces VG. Setting the boundaries of chromatin domains and nuclear organization. *Cell.* 2002;111(2):151–4.
3. Jenuwein T, Allis CD. Translating the histone code. *Science.* 2001;293(5532):1074–80.
4. Peterson CL, Laniel M-A. Histones and histone modifications. *Curr Biol.* 2004;14(14):R546–51.
5. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome resolution core particle at 2.8 Å. *Nature.* 1997;389:251–60.
6. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16(1):6–21.
7. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature.* 2000;403(6765):41–5.
8. Choy JS, Wei S, Lee JY, Tan S, Chu S, Lee TH. DNA methylation increases nucleosome compaction and rigidity. *J Am Chem Soc.* 2010;132(6):1782–3.
9. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, et al. Active genes are tri-methylated at K4 of histone H3. *Nature.* 2002;419(6905):407–11.
10. Sullivan BA, Karpen GH. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol.* 2004;11(11):1076–83.
11. Grewal SIS, Jia S. Heterochromatin revisited. *Nat Rev Genet.* 2007;8(1):35–46.
12. Berger SL. The complex language of chromatin regulation during transcription. *Nature.* 2007;447(7143):407–12.
13. Xiao B, Wilson JR, Gambelin SJ. SET domains and histone methylation. *Curr Opin Struct Biol.* 2003;13(6):699–705.
14. Pfannenstiel BT, Keller NP. On top of biosynthetic gene clusters: How epigenetic

- 754 machinery influences secondary metabolism in fungi. *Biotechnol Adv.*
755 2019;37(6):107345.
- 756 15. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine
757 methylations. *Exp Mol Med.* 2017;49(4).
- 758 16. Collemare J, Seidl MF. Chromatin-dependent regulation of secondary metabolite
759 biosynthesis in fungi: is the picture complete? *FEMS Microbiol Rev.* 2019;43(6):591–
760 607.
- 761 17. Helmlinger D, Papai G, Devys D, Tora L. What do the structures of GCN5-containing
762 complexes teach us about their function? *Biochim Biophys Acta - Gene Regul Mech.*
763 2021;1864(2).
- 764 18. Hajheidari M, Koncz C, Bucher M. Chromatin evolution-key innovations underpinning
765 morphological complexity. *Front Plant Sci.* 2019;10:1–12.
- 766 19. Gates LA, Foulds CE, O'Malley BW. Histone marks in the 'drivers seat': functional
767 roles in steering the transcription cycle. 2018;42(12):977–89.
- 768 20. Jamieson K, Wiles ET, McNaught KJ, Sidoli S, Leggett N, Shao Y, et al. Loss of HP1
769 causes depletion of H3K27me3 from facultative heterochromatin and gain of
770 H3K27me2 at constitutive heterochromatin. *Genome Res.* 2016;26(1):97–107.
- 771 21. Jamieson K, Rountree MR, Lewis ZA, Stajich JE, Selker EU. Regional control of
772 histone H3 lysine 27 methylation in *Neurospora*. *Proc Natl Acad Sci U S A.*
773 2013;110(15):6027–32.
- 774 22. Freitag M. Histone Methylation by SET Domain Proteins in Fungi. *Annu Rev*
775 *Microbiol.* 2017;71:413–39.
- 776 23. Kamei M, Ameri AJ, Ferraro AR, Bar-peled Y, Zhao F. IMITATION SWITCH is
777 required for normal chromatin structure and gene repression in PRC2 target domains.
778 2021;118(4).
- 779 24. Weiner AKM, Cerón-Romero MA, Yan Y, Katz LA. Phylogenomics of the epigenetic
780 toolkit reveals punctate retention of genes across eukaryotes. *Genome Biol Evol.*
781 2020;1–39.

25. Choi J, Kim S-H. A genome tree of life for the fungi kingdom. *Proc Natl Acad Sci*. 2017;114(35):9391–6.
26. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Grigoriev I V., James TY, et al. Diversity of cytosine methylation across the fungal tree of life. *Nat Ecol Evol*. 2019;3(3):479–90.
27. Kouzminova E, Selker EU. dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J*. 2001;20(15):4309–23.
28. Nai YS, Huang YC, Yen MR, Chen PY. Diversity of Fungal DNA Methyltransferases and Their Association With DNA Methylation Patterns. *Front Microbiol*. 2021;11.
29. Samson RA, Visagie CM, Houbraken J, Hong SB, Hubka V, Klaassen CHW, et al. Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Stud Mycol*. 2014;78(1):141–73.
30. Cánovas D, Marcos AT, Gacek A, Ramos MS, Gutiérrez G, Reyes-Domínguez Y, et al. The histone acetyltransferase GcnE (GCN5) plays a central role in the regulation of *Aspergillus* asexual development. *Genetics*. 2014;197(4):1175–89.
31. Palmer JM, Perrin RM, Dagenais TRT, Keller NP. H3K9 methylation regulates growth and development in *Aspergillus fumigatus*. *Eukaryot Cell*. 2008;7(12):2052–60.
32. Yang K, Liang L, Ran F, Liu Y, Li Z, Lan H, et al. The DmtA methyltransferase contributes to *Aspergillus flavus* conidiation, sclerotial production, aflatoxin biosynthesis and virulence. *Sci Rep*. 2016;6:1–13.
33. Palmer JM, Bok JW, Lee S, Dagenais TRT, Andes DR, Kontoyiannis DP, et al. Loss of CclA, required for histone 3 lysine 4 methylation, decreases growth but increases secondary metabolite production in *Aspergillus fumigatus*. *PeerJ*. 2013;2013(1):1–18.
34. Georgakopoulos P, Lockington RA, Kelly JM. The Spt-Ada-Gcn5 Acetyltransferase (SAGA) Complex in *Aspergillus nidulans*. *PLoS One*. 2013;8(6).
35. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7(10):e1002195.
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:

- improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
37. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3.
38. Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 2016;65(6):997–1008.
39. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
40. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 2011;60(5):685–99.
41. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019;47(W1):W256–9.
42. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2.
43. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, et al. BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* 2013;41(W1):W29–33.
44. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005;6(1):31.
45. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Elsevier current trends*; 2000.
46. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng.* 2007;9(03):90–5.
47. Waskom ML. Seaborn: statistical data visualization. *J Open Source Softw.* 2021;6(60):3021.
48. Noberini R, Restellini C, Savoia EO, Bonaldi T. Enrichment of histones from patient

- 838 samples for mass spectrometry-based analysis of post-translational modifications.
- 839 Methods. 2020;184:19–28.
- 840 49. He F. Bradford protein assay. Bio-protocol. 2011;45.
- 841 50. Noberini R, Savoia EO, Brandini S, Greco F, Marra F, Bertalot G, et al. Spatial epi-
842 proteomics enabled by histone post-translational modification analysis from low-
843 abundance clinical samples. Clin Epigenetics. 2021;13(1):1–16.
- 844 51. Prudhomme N, Gianetto-Hill C, Pastora R, Cheung W-F, Allen-Vercoe E, McLean
845 MD, et al. Quantitative proteomic profiling of shake flask versus bioreactor growth
846 reveals distinct responses of *Agrobacterium tumefaciens* for preparation in molecular
847 pharming. Can J Microbiol. 2021;67(1):75–84.
- 848 52. Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. Mass
849 spectrometry-based identification and characterisation of lysine and arginine
850 methylation in the human proteome. Mol Biosyst. 2013;9(9):2231–47.
- 851 53. Yuan Z-F, Sidoli S, Marchione DM, Simithy J, Janssen KA, Szurgot MR, et al.
852 EpiProfile 2.0: a computational platform for processing epi-proteomics mass
853 spectrometry data. J Proteome Res. 2018;17(7):2533–41.
- 854 54. Pesavento JJ, Mizzen CA, Kelleher NL. Quantitative analysis of modified proteins and
855 their positional isomers by tandem mass spectrometry: human histone H4. Anal
856 Chem. 2006;78(13):4271–80.
- 857 55. Wade PA, Pruss D, Wolffe AP. Histone acetylation: Chromatin in action. Trends
858 Biochem Sci. 1997;22(4):128–32.
- 859 56. Binda O. On your histone mark, SET, methylate! Epigenetics. 2013;8(5):457–63.
- 860 57. Wolffe AP, Matzke MA. Epigenetics: Regulation through repression. Science (80-).
861 1999;286(5439):481–6.
- 862 58. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a Fully Resolved Fungal Tree of
863 Life. Annu Rev Microbiol. 2020;74(1):291–313.
- 864 59. Sun B Da, Houbraken J, Frisvad JC, Jiang XZ, Chen AJ, Samson RA. New species in
865 *Aspergillus* section *Usti* and an overview of *Aspergillus* section *Cavernicolarum*. Int J

- Syst Evol Microbiol. 2020;
60. Chen AJ, Frisvad JC, Sun BD, Varga J, Kocsubé S, Dijksterhuis J, et al. *Aspergillus* section *Nidulantes* (formerly *Emericella*): Polyphasic taxonomy, chemistry and biology. Stud Mycol. 2016;84:1–118.
61. Steenwyk JL, Shen XX, Lind AL, Goldman GH, Rokas A. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. MBio. 2019;10(4):1–25.
62. Vesth TC, Nybo JL, Theobald S, Frisvad JC, Larsen TO, Nielsen KF, et al. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. Nat Genet. 2018;50(12):1688–95.
63. Kjærboelling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, et al. A comparative genomics study of 23 *Aspergillus* species from section *Flavi*. Nat Commun. 2020;11(1).
64. Souza SC, Pereira VM, Moreira SI, Costa SS, Moreira GM, Mendes WO, et al. *Aspergillus trisporus*: A new *Jani* section species from Brazilian soil. Curr Res Environ Appl Mycol. 2019;9(1):175–86.
65. Ko YH, So KK, Chun J, Kim DH. Distinct roles of two dna methyltransferases from *Cryptosporidium parvum* in fungal virulence, responses to hypovirus infection, and viral clearance. MBio. 2021;12(1):1–16.
66. Jurkowski TP, Jeltsch A. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One. 2011;6(11):1–9.
67. Bestor T. Structure of mammalian DNA methyltransferase as deduced from the inferred amino acid sequence and direct studies of the protein. Portland Press Ltd.; 1988.
68. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19(3):219–20.
69. Zhang Z, He C, Chen Y, Li B, Tian S. DNA methyltransferases regulate pathogenicity of *Botrytis cinerea* to horticultural crops. J Fungi. 2021;7(8).

70. Moller M, Habig M, Lorrain C, Feurtey A, Haueisen J, Fagundes WC, et al. Recent loss of the Dim2 DNA methyltransferase decreases mutation rate in repeats and changes evolutionary trajectory in a fungal pathogen. *PLoS Genet.* 2021;17(3):1–27.
71. Delmas V, Stokes DG, Perry RP. A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proc Natl Acad Sci.* 1993;90(6):2414–8.
72. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal Chem.* 2014;86(8):3697–702.
73. Huff JT, Zilberman D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell.* 2014;156(6):1286–97.
74. Proffitt JH, Davie JR, Swinton D, Hattman S. 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. *Mol Cell Biol.* 1984;4(5):985–8.
75. Catania S, Dumesic PA, Pimentel H, Nasif A, Stoddard CI, Burke JE, et al. Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a De Novo Methyltransferase. *Cell.* 2020;180(2):263-277.e20.
76. Rountree MR, Selker EU. DNA methylation and the formation of heterochromatin in *Neurospora crassa*. *Heredity (Edinb).* 2010;105(1):38–44.
77. Grognet P, Timpano H, Carlier F, Aït-Benkhalil J, Berteaux-Lecellier V, Debuchy R, et al. A RID-like putative cytosine methyltransferase homologue controls sexual development in the fungus *Podospora anserina*. *PLoS Genet.* 2019;15(8):e1008086.
78. Madhani HD. Unbelievable but True: Epigenetics and Chromatin in Fungi. *Trends Genet.* 2020;
79. Liu S-Y, Lin J-Q, Wu H-L, Wang C-C, Huang S-J, Luo Y-F, et al. Bisulfite sequencing reveals that *Aspergillus flavus* holds a hollow in DNA methylation. *PLoS One.* 2012;7(1):e30349.
80. Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and function of

- SET and MYND domain-containing proteins. *Int J Mol Sci.* 2015;16(1):1406–8.
81. Raunser S, Magnani R, Huang Z, Houtz RL, Trievel RC, Penczek PA, et al. Rubisco in complex with Rubisco large subunit methyltransferase. *Proc Natl Acad Sci U S A.* 2009;106(9):3160–5.
82. Trievel RC, Flynn EM, Houtz RL, Hurley JH. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSM1. *Nat Struct Biol.* 2003;10(7):545–52.
83. Porras-Yakushi TR, Whitelegge JP, Clarke S. Yeast ribosomal/cytochrome c SET domain methyltransferase subfamily: Identification of Rpl23ab methylation sites and recognition motifs. *J Biol Chem.* 2007;282(17):12368–76.
84. Roguev A, Schaft D, Shevchenko A, Pijnappel WWMP, Wilm M, Aasland R, et al. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* 2001;20(24):7137–48.
85. Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, et al. Structure of the *Neurospora* SET Domain Protein DIM-5, a Histone H3 Lysine Methyltransferase (Boggs et al methylation has been implicated in transcriptional silencing of eu. *Cell.* 2002;111:9.
86. Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD. A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev.* 2013;27(17):1851–6.
87. Strahl BD, Grant PA, Briggs SD, Sun Z-W, Bone JR, Caldwell JA, et al. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol.* 2002;22(5):1298–306.
88. Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, et al. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* 2003;17(5):654–63.
89. Li M, Phatnani HP, Guan Z, Sage H, Greenleaf AL, Zhou P. Solution structure of the Set2–Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proc Natl Acad Sci.*

- 2005;102(49):17636–41.
90. Bicocca VT, Ormsby T, Adhvaryu KK, Honda S, Selker EU. ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. *Elife*. 2018;7:1–19.
91. Janevska S, Baumann L, Sieber CMK, Münsterkötter M, Ulrich J, Kämper J, et al. Elucidation of the Two H3K36me3 Histone Methyltransferases Set2 and Ash1 in *Fusarium fujikuroi* Unravels Their Different Chromosomal Targets and a Major Impact of Ash1 on Genome Stability. *Genetics*. 2018;208:153–71.
92. Wang Y, Reddy B, Thompson J, Wang H, Noma K, Yates III JR, et al. Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol Cell*. 2009;33(4):428–37.
93. Veerappan CS, Avramova Z, Moriyama EN. Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. *BMC Evol Biol*. 2008;8(1):1–20.
94. Gacek-Matthews A, Noble LM, Gruber C, Berger H, Sulyok M, Marcos AT, et al. KdmA, a histone H3 demethylase with bipartite function, differentially regulates primary and secondary metabolism in *Aspergillus nidulans*. *Mol Microbiol*. 2015;96(4):839–60.
95. Bachleitner S, Sørensen JL, Gacek-Matthews A, Sulyok M, Studt L, Strauss J. Evidence of a Demethylase-Independent Role for the H3K4-Specific Histone Demethylases in *Aspergillus nidulans* and *Fusarium graminearum* Secondary Metabolism. *Front Microbiol*. 2019;10.
96. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone acetyltransferase complexes. *Trends Genet*. 2003;19(6):321–9.
97. Torchia J, Glass C, Rosenfeld MG. Co-activators and co-repressors in the integration of transcriptional responses. *Curr Opin Cell Biol*. 1998;10(3):373–83.
98. Sterner DE, Berger SL. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*. 2000;64(2):435–59.
99. Roth SY, Denu JM, Allis CD. Histone acetyltransferases. *Annu Rev Biochem*.

978 2001;70(1):81–120.

979 100. Kawahara T, Siegel TN, Ingram AK, Alsford S, Cross GAM, Horn D. Two essential
980 MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric
981 silencing in trypanosomes. *Mol Microbiol*. 2008;69(4):1054–68.

982 101. Takechi S, Nakayama T. Sas3 is a histone acetyltransferase and requires a zinc
983 finger motif. *Biochem Biophys Res Commun*. 1999;266(2):405–10.

984 102. Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, et al. Yng1 PHD Finger
985 Binding to H3 Trimethylated at K4 Promotes NuA3 HAT Activity at K14 of H3 and
986 Transcription at a Subset of Targeted ORFs. *Mol Cell*. 2006;24(5):785–96.

987 103. Gómez EB, Espinosa JM, Forsburg SL. *Schizosaccharomyces pombe* mst2+
988 Encodes a MYST Family Histone Acetyltransferase That Negatively Regulates
989 Telomere Silencing. *Mol Cell Biol*. 2005;25(20):8887–903.

990 104. Nugent RL, Johnsson A, Fleharty B, Gogol M, Xue-Franzén Y, Seidel C, et al.
991 Expression profiling of *S. pombe* acetyltransferase mutants identifies redundant
992 pathways of gene regulation. *BMC Genomics*. 2010;11(1):59.

993 105. Kim J, Park S, Lee JS. Epigenetic control of oxidative stresses by histone
994 acetyltransferases in *Candida albicans*. *J Microbiol Biotechnol*. 2018;28(2):181–9.

995 106. Millar CB, Xu F, Zhang K, Grunstein M. Acetylation of H2AZ Lys 14 is associated with
996 genome-wide gene activity in yeast. *Genes Dev*. 2006;20(6):711–22.

997 107. Searle NE, Torres-Machorro AL, Pillus L. Chromatin regulation by the NuA4
998 acetyltransferase complex is mediated by essential interactions between enhancer of
999 polycomb (Epl1) and Esa1. *Genetics*. 2017;205(3):1125–37.

1000 108. Allard S, Utley RT, Savard J, Clarke A, Grant P, Brandl CJ, et al. NuA4, an essential
1001 transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the
1002 ATM-related cofactor Tra1p. *EMBO J*. 1999;18(18):5108–19.

1003 109. Wang X, Chang P, Ding J, Chen J. Distinct and redundant roles of the two MYST
1004 histone acetyltransferases Esa1 and Sas2 in cell growth and morphogenesis of
1005 *Candida albicans*. *Eukaryot Cell*. 2013;12(3):438–49.

110. Meijsing SH, Ehrenhofer-Murray AE. The silencing complex SAS-I links histone acetylation to the assembly of repressed chromatin by CAF-I and Asf1 in *Saccharomyces cerevisiae*. *Genes Dev.* 2001;15(23):3169–82.
111. Balasubramanian R, Pray-Grant MG, Selleck W, Grant PA, Tan S. Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation. *J Biol Chem.* 2002;277(10):7989–95.
112. Winkler GS, Kristjuhan A, Erdjument-Bromage H, Tempst P, Svejstrup JQ. Elongator is a histone H3 and H4 acetyltransferase important for normal histone acetylation levels in vivo. *Proc Natl Acad Sci U S A.* 2002;99(6):3517–22.
113. Krogan NJ, Greenblatt JF. Characterization of a Six-Subunit Holo-Elongator Complex Required for the Regulated Expression of a Group of Genes in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 2001;21(23):8203–12.
114. Wittschieben B, Otero G, De Bizemont T, Fellows J, Erdjument-Bromage H, Ohba R, et al. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol Cell.* 1999;4(1):123–8.
115. Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y. A role for the elongator complex in zygotic paternal genome demethylation. *Nature.* 2010;463(7280):554–8.
116. Booker SJ, Grove TL. Mechanistic and functional versatility of radical SAM enzymes. *F1000 Biol Rep.* 2010;2(1).
117. Yuan Y, Yuan H, Yang G, Yun H, Zhao M, Liu Z, et al. IFN- α confers epigenetic regulation of HBV cccDNA minichromosome by modulating GCN5-mediated succinylation of histone H3K79 to clear HBV cccDNA. *Clin Epigenetics.* 2020;12(1):1–16.
118. Kollenstart L, De Groot AJL, Janssen GMC, Cheng X, Vreeken K, Martino F, et al. Gcn5 and esa1 function as histone crotonyltransferases to regulate crotonylation-dependent transcription. *J Biol Chem.* 2019;294(52):20122–34.
119. Jones R. Identification and Verification of Lysine Propionylation and Butyrylation in Yeast Core Histones Using PTMap Software. *Bone.* 2014;23(1):1–7.

1034 120. Liu B, Lin Y, Darwanto A, Song X, Xu G, Zhang K. Identification and characterization
1035 of propionylation at histone H3 lysine 23 in mammalian cells. J Biol Chem.
1036 2009;284(47):32288–95.

1037 121. Papanicolaou K, O'Rourke B, Foster DB. Metabolism leaves its mark on the
1038 powerhouse: Recent progress in post-translational modifications of lysine in
1039 mitochondria. Front Physiol. 2014;5 JUL.

1040 122. Ekwall K. Genome-wide analysis of HDAC function. Trends Genet. 2005;21(11):608–
1041 15.

1042 123. Honda S, Bicocca VT, Gessaman JD, Rountree MR, Yokoyama A, Yu EY, et al. Dual
1043 chromatin recognition by the histone deacetylase complex HCHC is required for
1044 proper DNA methylation in *Neurospora crassa*. Proc Natl Acad Sci U S A.
1045 2016;113(41):E6135–44.

1046 124. Job G, Brugger C, Xu T, Lowe BR, Pfister Y, Qu C, et al. SHREC silences
1047 heterochromatin via distinct remodeling and deacetylation modules. Mol Cell.
1048 2016;62(2):207–21.

1049 125. Wu J, Suka N, Carlson M, Grunstein M. TUP1 utilizes histone H3/H2B-specific HDA1
1050 deacetylase to repress gene activity in yeast. Mol Cell. 2001;7(1):117–26.

1051 126. Kurdistani SK, Grunstein M. Histone acetylation and deacetylation in yeast. Nat Rev
1052 Mol cell Biol. 2003;4(4):276–84.

1053 127. Li X, Pan L, Wang B, Pan L. The histone deacetylases HosA and HdaA affect the
1054 phenotype and transcriptomic and metabolic profiles of *Aspergillus Niger*. Toxins
1055 (Basel). 2019;11(9).

1056 128. Lee I, Oh JH, Keats Shwab E, Dagenais TRT, Andes D, Keller NP. HdaA, a class 2
1057 histone deacetylase of *Aspergillus fumigatus*, affects germination and secondary
1058 metabolite production. Fungal Genet Biol. 2009;46(10):782–90.

1059 129. Brosch G, Loidl P, Graessle S. Histone modifications and chromatin dynamics: A
1060 focus on filamentous fungi. FEMS Microbiol Rev. 2008;32(3):409–39.

1061 130. Carmen AA, Griffin PR, Calaycay JR, Rundlett SE, Suka Y, Grunstein M. Yeast HOS3

- forms a novel trichostatin A-insensitive homodimer with intrinsic histone deacetylase activity. *Proc Natl Acad Sci U S A*. 1999;96(22):12356–61.
131. Pérez-Martínez ME, Benet M, Alepuz P, Tordera V. Nut1/Hos1 and Sas2/Rpd3 control the H3 acetylation of two different sets of osmotic stress-induced genes. *Epigenetics*. 2020;15(3):251–71.
132. Li S, Yue Z, Tanaka TU. Smc3 Deacetylation by Hos1 Facilitates Efficient Dissolution of Sister Chromatid Cohesion during Early Anaphase. *Mol Cell*. 2017;68(3):605-614.e4.
133. Wang A, Kurdistani SK, Grunstein M. Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science*. 2002;298(5597):1412–4.
134. Suka N, Suka Y, Carmen AA, Wu J, Grunstein M. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol Cell*. 2001;8(2):473–9.
135. De Nadal E, Zapater M, Alepuz PM, Sumoy L, Mas G, Posas F. The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoreponsive genes. *Nature*. 2004;427(6972):370–4.
136. Tribus M, Bauer I, Galehr J, Rieser G, Trojer P, Brosch G, et al. A novel motif in fungal class 1 histone deacetylases is essential for growth and development of *Aspergillus*. *Mol Biol Cell*. 2010;21(2):345–53.
137. Kawauchi M, Nishiura M, Iwashita K. Fungus-specific sirtuin HstD coordinates secondary metabolism and development through control of LaeA. *Eukaryot Cell*. 2013;12(8):1087–96.
138. Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, et al. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem*. 2006;281(46):35404–12.
139. Lewis ZA, Adhvaryu KK, Honda S, Shiver AL, Knip M, Sack R, et al. DNA methylation and normal chromosome behavior in neurospora depend on five components of a histone methyltransferase complex, DCDC. *PLoS Genet*. 2010;6(11):1–11.

1090 140. Eberharther A, Sterner DE, Schieltz D, Hassan A, Yates JR, Berger SL, et al. The ADA
1091 Complex Is a Distinct Histone Acetyltransferase Complex in *Saccharomyces*
1092 *cerevisiae*. Mol Cell Biol. 1999;19(10):6621–31.

1093 141. Doyon Y, Côté J. The highly conserved and multifunctional NuA4 HAT complex. Curr
1094 Opin Genet Dev. 2004;14(2):147–54.

1095 142. Verzijlbergen KF, van Welsem T, Sie D, Lenstra TL, Turner DJ, Holstege FCP, et al.
1096 A barcode screen for epigenetic regulators reveals a role for the NuB4/HAT-B histone
1097 acetyltransferase complex in histone turnover. PLoS Genet. 2011;7(10):18–24.

1098 143. Sendra R, Tse C, Hansen JC. The yeast histone acetyltransferase A2 complex, but
1099 not free Gcn5p, binds stably to nucleosomal arrays. J Biol Chem.
1100 2000;275(32):24928–34.

1101 144. Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, et al. Yeast Gcn5
1102 functions in two multisubunit complexes to acetylate nucleosomal histones:
1103 Characterization of an ada complex and the saga (spt/ada) complex. Genes Dev.
1104 1997;11(13):1640–50.

1105 145. Steenwyk JL, Mead ME, Knowles SL, Raja HA, Roberts CD, Bader O, et al. Variation
1106 among biosynthetic gene clusters, secondary metabolite profiles, and cards of
1107 virulence across *Aspergillus* species. Genetics. 2020;216(2):481–97.

1108
1109

Figure captions

Figure 1. Conservation of DNA methyltransferase families in the *Aspergillus* genus. The maximum-likelihood phylogeny of DNA methyltransferase domains (PF00145) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (38,39). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in DNA methyltransferase domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Figure 2. Conservation of histone methyltransferase families in the *Aspergillus* genus. The maximum-likelihood phylogeny of SET domains (PF00856) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (38,39). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles and leaves with similar protein architecture with unknown functions covering both *Aspergilli* and outgroup species are collapsed and shown as grey triangles. Additional conserved domains in SET domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Figure 3. Conservation of MOZ/SAS FAMILY domain-containing histone acetyltransferases in the *Aspergillus* genus. The maximum-likelihood phylogeny of MOZ/SAS FAMILY domains (PF01853) found in 94 *Aspergilli* and 15 outgroup species was determined using IQ-TREE (38,39). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from *Aspergilli* with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in MOZ/SAS FAMILY domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Figure 4. Conservation of histone deacetylase domain-containing proteins in the *Aspergillus* genus. The maximum-likelihood phylogeny of Histone deacetylase domains

(PF00850) found in 94 Aspergilli and 15 outgroup species was determined using IQ-TREE (38,39). The black dot on the branch indicates ultrafast bootstrap values over 95 and SH-aLRT values over 80. Leaves from Aspergilli with trustworthy support are collapsed and shown as red triangles. Additional conserved domains in histone deacetylase domain-containing proteins were identified using the PFAM database (<https://pfam.xfam.org>).

Figure 5. Presence-absence pattern of subunits that constitute characterized histone modification complexes. The top row with distinct color boxes indicates outgroup species and Aspergilli sections, the order of the species from right to left is the same as in the species tree from top to bottom (Supplementary Figure 1) without *Fusarium oxysporum* and *Aspergillus kawachii* since they were excluded in the later analysis. Each row represents a subunit in the histone modification complexes, the order agrees with the summary list (Supplementary Table 3). Each box in the matrix indicates the presence (dark color) or absence (light color) of a subunit (Supplementary Table 3). Each row shows the presence/absence of a specific complex subunit. For each complex, the first row represents the catalytic subunit.

Figure 6. Schematic representation of the histone code in Aspergilli. (A) Relative abundance of post-translational modifications detected on peptides of histone H3 and H4 from *Aspergillus nidulans* using unbiased mass spectrometry analysis. The term “combinations” refers to multiple modifications occurring on the same peptide (Supplementary Table 4). **(B)** Histone modifications are predicted by genomic and phylogenetic analyses, and/or proved by mass spectrometry in *A. nidulans*. Histone modifications verified by mass spectrometry are labeled as solid dots, while the ones that are only predicted by phylogenic analysis are labeled as solid dots with a red outline.

1164 **Supplementary material**

1165 **Supplementary Table S1. List of species included in this study.**

1166 **Supplementary Table S2. List of BUSCO genes used to build the species tree.**

1167 **Supplementary Table S3. List of chromatin modifier complexes reported in literature.**

1168 **Supplementary Table S4. Mass spectrometry result for *A. nidulans*.**

1169

1170 **Supplementary Figure 1. A robust species phylogeny for the fungal genus *Aspergillus*.**

1171 **Supplementary Figure 2. Two known histone acyltransferase-domain containing protein**
 1172 **groups are conserved in the *Aspergillus* genus together with many unknown proteins.**

1173

1174 **Supplementary Data 1. Fasta, trimmed alignment, and maximum-likelihood phylogenetic**
 1175 **tree files associated with Figure 1-4.**

1176 **Supplementary Data 2. Fasta, trimmed alignment, and maximum-likelihood phylogenetic**
 1177 **tree files associated with Figure 5.**

1178

Tree scale: 1
 Histone deacetylase PF00850
 Arb2 PF09757

A **B**